DOCUMENT RESUME

ED 105 046 UD 015 037

AUTHOR	Samuel, William; And Others
TITLE	Motivation, Race, Social Class, and I.Q.
INSTITUTION	California State Univ., Sacramento. Dept. of Psychology.
SPONS AGENCY	National Inst. of Mental Health (DHEW), Rockville, Md.
PUB DATE	May 74
NOTE	41p.; Parts of this paper were presented at meetings of the Western Psychological Association (San Francisco, California, May 1974)
EDRS PRICE	MF-$0.76 HC-$1.95 PLUS POSTAGE
DESCRIPTORS	Anxiety; Individual Characteristics; Individual Differences; *Intelligence Differences; Intelligence Tests; *Motivation; Performance Factors; Psychological Characteristics; *Racial Differences; Sex Differences; Social Class; *Social Differences; *Social Psychology; Testing Problems
IDENTIFIERS	Wechsler Intelligence Scale for Children

ABSTRACT

While debates over the heritability of IQ and the potential for culture bias in measuring instruments have generated much research and public comment, it is also possible to investigate the significance of interracial differences in mean IQ by ignoring both the foregoing issues and instead examining the social psychology of the test situation itself. Male and female students between 12 and 16 years of age completed the Wechsler Intelligence Scale for Children Performance sub-scales in a variety of settings. The variables of test atmosphere (evaluative or gamelike), tester expectation (high or low), race of tester (black or white), and race of subject were placed in a two by two by two by two factorial design. At a second session some weeks after taking the WISC, subjects completed a group administered questionnaire. The pattern of mean IQ scores as well as mood and personality data indicated that test performance was optimal at moderate levels of motivational arousal. A replication of the experiment for male subjects increased cell sizes to the point that socio-economic status could be treated as an independent variable in the design. When this was done, the results suggested that interracial differences in mean IQ might be erased depending upon the social-psychological characteristics of the test setting and the socio-economic background of the testee. (Author/JM)

ED105046

Motivation, Race, Social Class, and IQ[1]

William Samuel[2], David Soto, Michael Parks,

Peter Ngissah[3], and Benjamin Jones

California State University, Sacramento

Running Title

Motivation and IQ

Mailing Address

Dr. William Samuel, Department of Psychology, California State University, Sacramento 95819.

Footnotes

1. This research was supported by a grant from the Public Health Service to the first author (MH 21285). The authors express their appreciation to Margrethe Cameron, who scored the WISC forms and recorded the data. Parts of this paper were presented at meetings of the Western Psychological Association, San Francisco, May 1974.

2. Requests for reprints should be sent to William Samuel, Department of Psychology, California State University, Sacramento 95819.

3. Now at the University of California, Davis.

UD015037

ABSTRACT

Male and female students between 12 and 16 years of age completed the WISC performance subscales in a variety of settings. The variables of test atmosphere (evaluative or gamelike), tester expectation (high or low), race of tester (black or white), and race of subject were placed in a 2x2x2x2 factorial design. The pattern of mean IQ scores as well as mood and personality data indicated that test performance was optimal at moderate levels of motivational arousal. A replication of the experiment for male subjects increased cell sizes to the point that socio-economic status could be treated as an independent variable in the design. When this was done, the results suggested that interracial differences in mean IQ might be erased depending upon the social-psychological characteristics of the test setting and the socio-economic background of the testee.

Date submitted: 3/10/75

Motivation, Race, Social Class, and IQ

If an IQ test is administered to children attending a racially integrated school, blacks will generally average from 11 to 15 points below whites (though there will be a substantial overlap in the two distributions). Jensen's (1969) suggested explanation for this phenomenon was that, since IQ has a very high heritability, large mean differences between racial groups must be predominantly genetic in origin. Environmentalists (e.g., Garcia, 1972) have replied that any such interracial differences are entirely attributable to an alleged bias in the content of IQ tests in favor of cultural experiences more readily accessible to middle and upper income whites.

While debates over the heritability of IQ and the potential for culture bias in measuring instruments have generated much research and public comment, it is also possible to investigate the significance of interracial differences in mean IQ by ignoring both the foregoing issues and instead examining the social psychology of the test situation itself. Sattler's (1970) review of the extensive literature on this topic found support for several hypotheses (among others): black children may have generally lower achievement motivation than whites; expectations of failure or fear of appearing "uppity" may impair the performance of black children when they anticipate comparison with whites; the performance of black children may be improved by providing a same-race tester; the performance of all children may be improved when the tester has a favorable rather than an unfavorable expectation of their ability [when overtly revealed, such expectations may result in what Rosenthal (1965) has called an "experimenter"

or instructional effect]. Since black students have in the past usually been tested by white examiners in what are often competitive, ego-threatening situations, it is conceivable--if black examinees tend to believe that practically any white tester will be prejudiced against them--that performance-debilitating self-fulfilling prophecies could be set in motion (Rosenthal and Jacobsen, 1968).

However, the influence of the test setting on the observed ability of black and white testees is not as consistently predictable as the preceding paragraph implies. Katz (1970), a prominent researcher in this area, has found, rather paradoxically, that the performance of black students often improves when they are tested by a white rather than by a black examiner, when they are in the presence of white rather than black agemates, or when they are to be compared against white rather than black norms. According to Katz, this may occur because whites are regarded as evaluators that one should try harder to impress or because comparison with white standards is more informative for self-evaluations of ability.

A study with hybrid results was recently reported by McClelland (1974). He found that both black and white subjects were more cooperative and motivated to achieve when a white rather than a black interviewer asked them to complete a battery of items from intelligence and personality tests. But both groups scored higher in intelligence in the presence of a black interviewer, perhaps, as McClelland suggests, because the white tester stimulated evaluation apprehension along with achievement motivation, thus producing "lower intelligence test scores, due to higher anxiety."

It may in fact be the case that the single construct of testee motivation (one element of which is test anxiety) could account for this complex array of evidence. Wine (1971) suggested that anxiety elicits "two classes of

responses: those related to task completion, which are anxiety reducing, and those which interfere with task completion." When anxiety is extremely high, the performer's state of internal arousal becomes disruptively distracting, which causes the interfering responses to predominate and which in turn debilitates performance. Sarason (1961) found that under conditions of ego threat habitually low test anxious subjects will surpass those whose performance has been debilitated by high test anxiety. Sarason also found, however, that under relaxed, non-threatening conditions, subjects low in test anxiety solved fewer anagrams than their habitually high test anxious counterparts; presumably, the latter were experiencing only a moderate, optimally motivating level of arousal due to the relaxed conditions whereas the former were not sufficiently motivated to complete the task. Weiner and Samuel (in press) administered Sarason's anagrams in an ego-threatening environment to chronically high test anxious subjects, some of whom were led to mislabel their test-induced physiological symptoms of anxiety as being due to the side effects of a capsule (placebo) which they had swallowed earlier. This group rated itself less anxious and was able to solve more anagrams than controls not given the opportunity to mislabel the source of their symptoms of internal arousal.

Such processes could account for the seemingly contradictory effects for race of tester on testee performance which have emerged in the literature. A white tester may be perceived as a more powerful evaluator than a black counterpart and so will elicit a better performance from testees so long as the greater internal arousal associated with his presence goes no higher than the optimal, moderate level. Certain environments may, however, induce some anxiety or arousal independently of the characteristics of the examiner. Under these conditions, the presence of a white evaluator could be excessively

arousing and so might result in performance impairment; instead, a less disruptively stimulating black evaluator might be able to elicit a superior performance.

In the present experiment, race of subject and race of tester were systematically varied. In addition, there were two test atmospheres such that students in one condition were explicitly told they were completing a battery of competitive tests while those in another condition thought they were working with a set of creative games and playthings. The tester also expressed either a high or a low expectation for the subject's probable performance. In all cells of this 2x2x2x2 factorial pairing of race of tester, race of subject, test atmosphere, and tester expectation, subjects completed the performance subscales to the Wechsler Intelligence Scale for Children.

It was anticipated that whites would, overall, score 11-15 points higher in IQ than blacks, since this appears to be a stable finding in the literature. For subjects of both races, however, it was predicted that in the evaluative atmosphere the ego-threatening and competitive nature of the instructions would induce a state of fairly high anxiety in testees and that if the tester then expressed a high expectation this might reduce anxiety slightly to a more moderate level and so facilitate performance; if the tester instead expressed a <u>low</u> expectation in the evaluative atmosphere, the added stress of his criticism should definitely debilitate performance. In the gamelike atmosphere, by contrast, the setting was anticipated to be so relaxed that testee motivation would be insufficient for optimal performance. Here it was predicted that achievement would be facilitated rather than debilitated by the moderate anxiety induced by a tester's low expectation.

It was further speculated that there might be situations in which

reactance motivation could override the debilitative effects of high test anxiety. Brehm's (1966) theory of reactance states that when people feel their freedom of action is being threatened by manipulation or coercion they will resist the threat and seek to emphasize their freedom to behave oppositely. In the present research, it was felt that if testees in the ego-threatening evaluative atmosphere were challenged by a low expectation on the part of an opposite-race tester, an especially strong desire to disprove the tester's negative assessment might lead to an effort at suppressing task-interfering responses to permit a resolute concentration on task completion. Baron and Ganz (1972) have suggested that reactance motivation might be especially likely to be aroused in black students confronted by a white evaluator, and Allen, Dubanoski, and Stevenson (1966) have reported that among older children criticism from a white experimenter was actually more effective than praise in maintaining the performance level of black testees.

To summarize, it was anticipated that in the evaluative atmosphere subjects would perform better on the WISC following an expressed high expectation on the part of the tester (an instructional effect). In the gamelike atmosphere, however, it was predicted that performance would be optimal following a tester's <u>low</u> expectation; in the sense that the subject was predicted to behave oppositely from the tester's overt "demand," this might be called a reactance effect, though other motivational states, such as anxiety or irritation resulting from the tester's criticism, were also expected to contribute to the phenomenon. For students in the evaluative atmosphere, it was speculated that an opposite-race tester's low expectation might be viewed as a challenge; if so, it might arouse an especially strong motivation to disconfirm the tester's negative assessment, leading

to a reactance effect rather than the instructional effect which was otherwise predicted for the evaluative atmosphere. Overall, it was anticipated that if performance responded as predicted to the manipulation of the testee's motivational state, the size of interracial differences in mean IQ would be found to be somewhat more flexible than was suggested by Jensen's (1969) review of the literature.

EXPERIMENT I

Method

During [illegible] the WISC performance measures were administered to 208 black a[illegible]e junior high and high school students between 12 and 16 years of age, equally divided by sex. The 2x2x2x2x2 factorial design varied test atmosphere (evaluative or gamelike), tester expectation (high or low), race of tester (black or white) and race and sex of subject.

Arriving at an office provided by the school, the subject encountered $Experimenter_1$ of a team of two male experimenters. One team consisted of two white and the other of two black experimenters. $Experimenter_1$ described himself as a representative of Psychology Incorporated, a company which manufactures either "tests of intelligence and mental capacity" (evaluative atmosphere) or "creative games and playthings" (gamelike atmosphere). In the evaluative condition the tester further declared that the subject's performance on the tests would be compared to that of other students at the school and against city and nationwide norms. In the gamelike condition, $Experimenter_1$ assured the subject that performance on the tasks was the subject's "own thing" and that he should relax and take it easy since "no one is going to be compared to anyone else here." To supplement these manipulations, the tester wore a tie and jacket in the evaluative conditions but removed the jacket and loosened the tie when the atmosphere was to be

gamelike. All experimenters were in their middle or late twenties.

Experimenter$_1$ revealed that Psychology Incorporated had reviewed the student's grades in "think" courses like math, English, or art as well as "action" courses like gym, shop, or home economics to arrive at a prediction for the subject's performance. Those in the high expectation conditions were told they could expect to have an easy time with the tasks while those in the low expectation treatments were told they would probably have a difficult time.

The student was then given the Object Assembly subtest of the WISC. In the evaluative atmosphere, an imposing interval timer was used to score performance on the items; in the gamelike atmosphere, the examiner used a wall clock and explained that he regretted having to time the activities but must try to keep things on schedule since other students would be arriving later. When the subtest was completed, Experimenter$_1$ reinforced the expectation manipulation by announcing that the subject had either done rather well, above the average, or rather poorly, below the average. He then explained that his partner, Mr. _____, had a few other tests (or games) for the student to work with. As he departed, Experimenter$_1$ removed the interval timer from the table in the evaluative atmosphere or put on his jacket and straightened his tie in the gamelike condition; he was then replaced by Experimenter$_2$, who was blind as to the subject's prior treatment. Experimenter$_2$ administered the Picture Arrangement, Picture Completion, Block Design, and Coding subtests.

After completing the WISC, the subject made a self-rating of performance on a 7-unit scale running from "Very poorly" to "Very well." Next, he or she filled out a mood adjective checklist (Nowlis, 1965). The adjectives on the checklist comprise scales for measuring aggression (e.g., "angry"),

anxiety ("clutched up"), concentration ("engaged in thought"), egotism ("boastful"), elation ("overjoyed"), fatigue ("tired"), sadness ("sorry"), skepticism ("suspicious"), and surgency ("playful"). For each adjective, the subject indicated the degree to which it described his or her feelings on a 4-unit scale running from "definitely not" to "definitely." Finally, subjects were asked their home address and the occupations of their parents. Parents' occupations were referred to the tables of ranked occupational categories in Duncan, Featherman, and Duncan (1972). The addresses were referred to census tracts as another measure of social class. The combined social class index ranged from 0 to 100.

At a second session some weeks later, subjects completed a group-administered questionnaire. This included a fully validated version of Rotter's internal-external scale adapted for use with children by Nowicki and Strickland (1973). According to Rotter (1966) "internals" generally believe they have control over the events which occur in their lives while "externals" believe their fates are decided by powerful deliberate or circumstantial forces beyond their control. Higher scores on the I-E scale are associated with greater externality. Also completed was the Marlowe-Crowne Social Desirability Scale (Crowne and Marlowe, 1960), which measures the strength of the subject's need for approval from others. In addition, two subscales for general and test-specific anxiety from Janis and Field (1959) were included on the questionnaire.

The last items on the questionnaire specifically tapped attitudes toward women and blacks. One item for each target group asked the subject to indicate the frequency with which he or she thought the group had encountered discrimination, on a scale running from "never" to "extremely often." A second item asked the degree to which the subject felt women

and blacks should oppose discrimination when and if it occurred, from "Always should relax and go along" to "Always stand up aggressively." The third item asked for the degree to which the subject agreed with whatever he or she felt was meant by "black is beautiful" or "women's liberation," respectively. Scores on the second and third items were combined for each target group into indices labeled "black is beautiful," for attitudes toward black assertiveness, and "women's liberation," for attitudes toward female assertiveness. Each of the three items was rated on a six-unit scale.

Since the subjects were students under 16 years old, especial care was taken to see that each participant left the first session in a pleasant frame of mind. Particularly in the low expectation treatments, the experimenters emphasized that they had really wanted to study the effect of a person's mood on his or her test performance and that in order to accomplish this it had been necessary for the tester to exaggerate some of the things he had said about their abilities. Subjects were reassured as to the quality and complete confidentiality of their own performance, paid $3.00, sworn to secrecy, and released. One index of subject satisfaction is the degree to which they maintained silence. A probe for prior knowledge was conducted both before and during each debriefing; it did not prove necessary to discard any subject for suspicion induced by a prior participant's breach of confidence.

RESULTS

Success of the Experimental Manipulations

Self-ratings of performance were considerably more positive for subjects in the high than in the low expectation conditions ($F(1,412) = 235.28$, $p < .001$). Responses to the mood checklist also tended to confirm

the success of the atmosphere as well as the expectation manipulations. Elation was greater in the high expectation treatment than in the low ($F(1,412) = 4.93$, $p < .05$) and was also greater in the gamelike than in the evaluative atmosphere ($F(1,412) = 19.95$, $p < .01$). Anxiety was greater in the evaluative than in the gamelike atmosphere ($F(1,412) = 7.88$, $p < .01$).

IQ Data

The IQ scores reported below are derived from the four subtests administered by Experimenter$_2$, prorated according to procedures in the WISC scoring manual (Wechsler, 1949). The prorating caused calculated IQs to be slightly higher than they would have been if the discarded Object Assembly score had instead been included and no prorating applied.

Table 1 and Figure 1 show the mean IQ measured in each cell of the experimental design (n = 13 subjects per cell). Male and female subjects

Insert Table 1 and Figure 1 about here

did not differ appreciably in IQ ($F(1,384) = 1.79$, n.s.), but white students scored higher in overall IQ than black students ($F(1,384) = 109.45$, $p < .001$). The overall mean IQ for whites was 111.13 while that for blacks was 96.67; the overall difference in mean IQ between the races was thus 14.46 points. Students of both races generally performed better in the presence of a white rather than a black tester ($F(1,384) = 17.11$, $p < .01$).

A significant Atmosphere x Expectation interaction ($F(1,384) = 6.50$, $p < .02$) developed from the mean IQs shown in Table 2. These means

Insert Table 2 about here

combined the scores of male or female, black or white subjects tested by black or white experimenters. In an evaluative atmosphere, students scored higher in IQ if they were told they would do well rather than poorly, but

this difference was not significant (t = 1.15). More specifically, however, white males in the evaluative atmosphere with a white tester scored 121.46 in mean IQ following a high expectation but only 110.31 if the tester's expectation was low, a significant instructional effect ($p < .05$). In the gamelike atmosphere, students did best when the tester was critical rather than encouraging, the predicted reactance effect (t = 2.47, $p < .002$).

An Atmosphere x Expectation x Sex of Subject x Race of Experimenter interaction was also observed ($F(1,384) = 2.80$, $p < .10$), but it was of only marginal reliability. While interpretation of a four-factor interaction is rather difficult, the pattern of mean IQs shown in Figure 1 is suggestive of the following: The Atmosphere x Expectation interaction was strongest in the presence of a black tester and was, overall, stronger for males than for females.

Correlates of IQ

Scores on the socio-economic index were positively correlated with IQ for both male (r = +.39; $t(414) = 6.11$, $p < .002$) and female (r = +.20; $t(414) = 2.98$, $p < .02$) subjects. In other words, students from more advantaged home environments tended to score higher in IQ.

Emotional and personality correlates of IQ are shown in Table 3. On the mood checklist, relaxed and happy mood states, like elation and

Insert Table 3 about here

surgency, were negatively related to IQ, as were tense emotional states like aggression or unhappy states like fatigue and sadness. Concentration was positively related to IQ, but only the data for whites were statistically significant. In general, though, these relationships were more often significant for blacks than for whites and for males than for females.

Subject self-ratings of performance were positively correlated with IQ, but only significantly so for whites. On the personality measures, general anxiety was negatively related to IQ for blacks but not significantly so for whites. Test anxiety and an "external" or fatalistic view of life on the I-E scale were negatively correlated with IQ for both blacks and whites.

Reactance Effects in the Evaluative Atmosphere

It was suggested in the introduction that examinees in the ego-threatening evaluative atmosphere who received a low expectation from an opposite-race tester might be motivated to apply themselves to task-completion so as to disprove the tester's negative assessment. If it occurred, such resistance would be manifested in peak performance following a low expectation in the evaluative atmosphere (a reactance effect) rather than the otherwise-predicted instructional effect. In Figure 1 it appears that the only reactance-type effects which were observed in the evaluative atmosphere occurred among black males and white females in the presence of a white tester. For white males and black females in the presence of a white tester and for all subjects in the presence of a black tester, peak performance was observed in the evaluative atmosphere following a high rather than a low tester expectation.

Thus, reactance-type effects occurred in the evaluative atmosphere only in the presence of a white tester. Moreover, as was explained earlier, a significant main effect for Race of Tester ($p < .01$) as well as a marginally significant four-factor interaction involving the race of the tester ($p < .10$) were disclosed in analyses of the IQ data. These findings suggested that the data gathered by white and black experimenters should be separated to permit a more detailed analysis. As

can be seen in the left half of Figure 1, a black examiner induced the same pattern of mean IQs whether he was working with black males, black females, white males, or white females: an instructional effect in the evaluative atmosphere and a reactance effect in the gamelike. The U-shaped curves are a graphic representation of the Atmosphere x Expectation interaction which was mentioned earlier; an analysis of variance on the data gathered by a black tester revealed this interaction in significant strength ($F(1,192) = 5.03$, $p < .05$).

IQs for white and black subjects faced with a white tester are shown in the right half of Figure 1. Perhaps the most striking feature of these data is the degree to which the curves for males and females intersect, indicating a rather opposite reaction on the part of the two sexes to the various test settings. Among blacks exposed to an evaluative atmosphere, females conformed to the white tester's expectations while males resisted this manipulation and did best when the tester forecast a poor performance. Among whites in the evaluative atmosphere it was males who conformed to the tester's expectations and females who resisted. The sharply contrasting reactions of male and female subjects to the expectation treatment thus had an additional racial component in that black males resisted while white males conformed, and white females resisted while black females conformed. Consequently, an analysis of variance revealed a significant Expectation x Sex of Subject x Race of Subject interaction ($F(1,192) = 4.61$, $p < .05$). In addition, a marginally reliable Atmosphere x Expectation x Sex of Subject interaction confirmed that these contrasting responses on the part of male and female subjects were most pronounced in the evaluative atmosphere ($F(1,192) = 3.40$, $p < .07$).

The only reactance effects observed in the evaluative atmosphere, then, were for black males in the presence of an opposite-race tester and white females in the presence of an opposite-sex but same-race tester. How might these phenomena be interpreted?

The overall positive correlation between black is beautiful and IQ which was found for black (r = +.22) but not for white (r = -.09) males in Table 3 may provide a clue as to the psychological processes underlying the reactance effect shown by black males in the presence of a white tester in the evaluative atmosphere. One finds that in this setting the correlation between black is beautiful and IQ became still more positive for black males (r = +.32). In the gamelike atmosphere with a white tester, by contrast, the correlation between black is beautiful and IQ was negative for black males (r = -.30). The difference between these correlations was significant (z = 2.13, $p < .02$). Belief in black is beautiful was particularly positively correlated with IQ for black males who received a low expectation from a white tester in the evaluative atmosphere (r = +.49; $t(11) = 1.90$, $p < .10$).

Although no overall relationship between women's liberation and IQ was found for female subjects of either race, the results for white females did parallel those for black males in certain respects. In the evaluative atmosphere with a white tester, women's liberation was positively correlated with IQ (r = +.21). In the gamelike atmosphere the relationship was negative (r = -.49; $t(24) = 2.76$, $p < .02$). The difference between these correlations was statistically significant (z = 2.55, $p < .01$). In the low expectation condition in the evaluative atmosphere with a white male tester the IQ of white females was positively related to belief in women's liberation but not significantly so (r = +.22).

Perhaps being in an evaluative atmosphere with a white male tester somehow stimulated the group pride of black males and white females, leading to an arousal of reactance motivation when they were challenged by a low expectation. If so, the results indicate that for these groups the reactance aroused by the tester's challenge was powerful enough to override the otherwise general tendency to perform better after receiving a high expectation in the evaluative atmosphere. There were no comparable findings in either the IQ or the personality data for white males or black females.

DISCUSSION

It appears that in the non-evaluative gamelike atmosphere test performance was facilitated rather than debilitated by the moderate anxiety or reactance motivation induced by an examiner's low expectation. In the evaluative atmosphere, by contrast, anxiety was by the nature of the experimental manipulations induced to be moderately high from the start; here, the added stress of an expressed low expectation on the part of the tester should have been debilitating. With the exception of the reactance effects found for black male and white female subjects in the evaluative atmosphere with a white tester, these predictions were substantially confirmed, as can be seen in Table 2.

The correlational data in Table 3 support the hypothesis that a moderate level of internal arousal induces optimal performance on an intellectual task. With less than moderate arousal, the performer will not be motivated to take the task seriously and so will focus insufficient attention on its completion. Thus, relaxed mood states like elation and surgency as well as depressive mood states like fatigue and sadness or a fatalistic, external world view were negatively related to IQ in Table 3. With more than moderate arousal, however, the performer will be distracted by his

internal state and may fail utterly. Thus, aggression in addition to general and test anxiety were negatively correlated with IQ in Table 3.

Even though observed IQ seems to have been reliably altered in response to the experimental manipulations, there was one major respect in which the data were disappointing: The flexibility of interracial differences in IQ was not convincingly demonstrated; there was virtually no overlap in mean IQ between the various groups of black and white subjects. A replication of the experiment did, however, demonstrate the anticipated manipulability of interracial IQ differences when socio-economic status was treated as an independent variable.

EXPERIMENT II

Method

The research was conducted during 1973-74 at two Sacramento junior high schools different from those used for Experiment I. The WISC performance measures were administered to 104 white and 104 black male students between 12 and 16 years of age. The variables of test atmosphere, tester expectation, race of tester, and race of subject were placed in a 2x2x2x2 factorial design. In all other respects, the procedure was identical to that utilized in Experiment I.

RESULTS

Since Experiment II duplicated procedures employed with male subjects in Experiment I, the two sets of data are discussed together in the analyses which follow. Hereafter, the results for males in Experiment I will be referred to as the 1972-73 experiment and the results for males in Experiment II as the 1973-74 experiment.

Success of the Manipulations

Across the 1972-73 and 1973-74 experiments, self-ratings of performance

were considerably more positive for subjects in the high than in the low expectation conditions ($F(1,412) = 273.74$, $p < .001$). Elation, too, was greater in the high expectation conditions than in the low ($F(1,412) = 14.36$, $p < .01$), while anxiety was greater in the evaluative than in the gamelike atmosphere ($F(1,412) = 5.74$, $p < .05$).

IQ Data

Shown in Figure 2 are the mean IQ scores from the 1972-73 and 1973-74 experiments (each point representing 13 subjects). The ways in which the

Insert Figure 2 about here

second experiment replicated the first will be considered before the relatively minor differences between these sets of data are discussed.

In both studies, whites scored higher in IQ than blacks ($F(1,384) = 79.59$, $p < .001$). The overall mean IQ for whites was 112.25 while that for blacks was 99.91, an interracial difference of 12.34 points. Students of both races performed better in the presence of a white rather than a black tester ($F(1,384) = 23.17$, $p < .01$).

Also in both experiments, a significant Atmosphere x Expectation interaction ($F(1,384) = 8.74$, $p < .01$) developed from the mean IQs shown in Table 4. These means combined the IQs of black or white male subjects

Insert Table 4 about here

tested by white or black experimenters. In an evaluative atmosphere, students scored higher in IQ if told they would do well than if told they would do poorly ($t = 2.24$, $p < .05$), an instructional effect. In the gamelike atmosphere, students did best when the tester was critical rather than encouraging ($t = 1.97$, $p < .05$), a reactance effect. The generally U-shaped curves in Figure 2 are the graphic representation of

the Atmosphere x Expectation interaction shown in Table 1.

The results for the 1973-74 experiment differed from those gathered for male subjects in 1972-73 in just two significant respects: First, subjects in the 1973-74 study had marginally higher IQs than those in the 1972-73 experiment ($F(1,384) = 3.30$, $p < .10$). Second, black subjects in the 1972-73 experiment who received a low expectation in the evaluative atmosphere from a white tester scored 5.00 points above black subjects given a high expectation in this setting. In 1973-74, however, black subjects who received a high expectation in the evaluative atmosphere from a white tester scored 7.92 points above those given a low expectation. Underlying both of the foregoing differences between the two experiments may be the fact that students in the 1973-74 study were of higher SES (12 points on the 100-unit scale) than those in the 1972-73 research ($t(414) = 7.85$, $p < .001$). The four schools from which students were sampled each had approximately the same proportion of black students (about 25%), but the two schools in which the 1973-74 experiment was conducted were located in more prosperous neighborhoods.

Consequently, the data for male subjects in the 1972-73 and 1973-74 studies were combined and the population divided into groups above and below the median in SES. The results are shown in Table 5 and Figure 3.

Insert Table 5 and Figure 3 about here

Clearly, subjects above the median in SES scored substantially higher in IQ than those below the median ($F(1,384) = 24.41$, $p < .01$).

All but one of the functions in Figure 2 is U-shaped, indicating that both high and low SES subjects displayed the Atmosphere x Expectation interaction mentioned earlier, with low SES blacks providing the sole

exception. The latter, after receiving a low expectation from a white tester in an evaluative atmosphere, scored 2.41 points above their high expectation counterparts. Though this is not a significant difference, it is the same phenomenon which was observed in the 1972-73 experiment and which failed to replicate in 1973-74, apparently because the latter population contained a greater proportion of high SES members.

Table 5 and Figure 3 indicate that high SES black students responded to the experimental manipulations in much the same way as did whites of either high or low SES. High SES black students did rather well on the WISC: When they were given encouragement by a white male tester in the evaluative atmosphere, their mean IQ reached 114.60, a value exceeded by whites in only three out of sixteen cells. Two interactions are relevant to this finding: Atmosphere x SES x Race of Experimenter ($F(1,384) = 4.74$, $p < .05$), which seems to have developed from the fact that within each racial group the best performance was recorded for high SES students in the presence of a white tester in the evaluative atmosphere, and Atmosphere x SES x Race of Subject x Race of Experimenter ($F(1,384) = 8.76$, $p < .01$), which is somewhat attributable to the observation that the IQ of high SES blacks equaled that of low SES whites in the evaluative atmosphere with a white tester and in the gamelike atmosphere with a black tester.

Mood, Motivation, and IQ

The personality and mood correlates of test performance for males in the 1972-73 and 1973-74 experiments will not be described at length, since they paralleled the findings shown in Table 3 for Experiment I. More directly relevant to the hypothesis that internal arousal must be at a moderate level for optimal performance are the mean scores for aggression and anxiety, shown in Table 6.

Insert Table 6 about here

Anxiety and aggressive motivation were minimal in the relaxed gamelike atmosphere when the tester praised the subject's abilities, but both arousal states showed an increase when the tester expressed a low expectation in this setting. As Table 4 indicates, IQ increased along with the increasing motivation. In the more stressful evaluative atmosphere, however, anxiety and aggression seem to have become excessive when the tester induced the subject's state of internal arousal to go beyond the optimal level through criticism of the latter's ability. Here, it was the encouragement offered by a high tester expectation which maintained arousal at a moderate level and permitted peak performance on the WISC.

Reactance Effects Among Low SES Students

It was noted earlier that low SES black students whose ability was criticized by a white tester in the evaluative atmosphere seemed to resist the tester's low expectation by outscoring their counterparts in the high expectation condition. In Experiment I, this phenomenon was observed among white females as well as black males, and it was suggested that the group pride of these subjects was challenged by a white male tester to a degree not felt by white male or black female subjects.

The present data indicate, however, that low SES white males may also to some extent be challenged by a white tester's low expectation. In a separate analysis of the IQ data gathered by a white tester (that is, the right half of Figure 3) a marginal Expectation x SES interaction emerged ($F(1,192) = 2.73$, $p = .10$). In general, high SES subjects performed better on the WISC after being encouraged by a high expectation while low SES subjects tended to do better following a low expectation. In addition,

an Atmosphere x SES interaction ($F(1,192) = 4.78$, $p < .05$) revealed that high SES subjects excelled in the evaluative atmosphere while low SES subjects performed best in the gamelike (especially, it appears in Figure 3, if the tester expressed a low expectation). Even if attention is restricted to the evaluative atmosphere, however, the Expectation x SES interaction persists ($F(1,96) = 3.01$, $p < .10$). Finally, of course, the Atmosphere x Expectation interaction was also found to be significant ($F(1,192) = 5.11$, $p < .05$). None of the foregoing effects interacted with the race of the subject (all Fs < 1).

When the IQ data gathered by a black tester (the left half of Figure 3) were separately analyzed, the Expectation x SES interaction did not appear ($F < 1$). An Atmosphere x SES x Race of Subject interaction ($F(1,192) = 4.80$, $p < .05$) reflected for the most part the equalization of high SES black and low SES white IQs in the gamelike atmosphere, and the Atmosphere x Expectation interaction was also significant ($F(1,192) = 4.17$, $p < .05$). Despite these reliable effects, however, it seems that a black tester did not motivate or challenge subjects to the same degree as his white counterpart; subjects of both races scored lower in IQ in the presence of a black tester. This could mean that a black tester was not taken as seriously as a white one (that is, subjects did not try as hard to impress him), so the changes in IQ induced by his communication of the atmosphere and expectation manipulations would have worked off a lower baseline of testee motivation. Black students seemed to be inspired to achieve a relatively high IQ in the presence of a black tester only when those of high SES were startled by a low expectation in what they had been led to believe was a "do your thing" gamelike atmosphere. Naturally, it must be kept in mind that just one black and one white experimenter gathered the IQ data. Any effects

attributed to race of tester are potentially confounded by the personalities of the individual experimenters and their proficiency in administering the WISC. Only further research and replication can clarify the mechanisms underlying race of tester effects.

DISCUSSION

Among both black and white subjects, instructional effects (peak performance in response to praise) predominated in the evaluative atmosphere while reactance effects (peak performance in response to criticism) predominated in the gamelike. This Atmosphere x Expectation interaction is interpreted as signifying that, in the ego-threatening evaluative atmosphere, internal arousal (one component of which is anxiety) was optimal when the subject was reassured by a high tester expectation but became excessive and, hence, performance-debilitating when the tester was critical. In the relaxed gamelike atmosphere, by contrast, a tester's low expectation served to elevate arousal from a low, insufficiently motivating level to a moderate, optimally motivating one and so facilitated test performance. In Tables 2 and 4 it can be seen that the range of variation in mean IQ which appears to be attributable to this Atmosphere x Expectation interaction is around 4-5 points.

For any testee, then, the most facilitative environment seems to be one which develops and maintains internal arousal at an optimal, moderate level, avoiding the extremes of anxiety or disinterest. Intriguingly, Doob and Kirshenbaum (1973), in a study of the effects of frustration and aggressive films on emotional arousal, similarly discovered that performance on a digit symbol task was a U-shaped function of arousal. Moderately elevated levels of blood pressure produced peak performance on the digit symbols while normal resting levels or excessively high levels served

to debilitate performance.

It is possible, of course, that some construct other than testee motivation might be able to account for the data. Since the experimenters overtly communicated their expectations to the subjects, demand effects were no doubt operative (Orne, 1962). While such demands could explain the instructional effects in the evaluative atmosphere, however, they cannot easily account for the reactance effects in the gamelike setting. Furthermore, $Experimenter_2$, who administered the subscales from which a given subject's IQ was calculated, was blind as to the subject's prior treatment by $Experimenter_1$. All experimenters were kept ignorant of the hypotheses until the conclusion of the research, but regardless of that precaution $Experimenter_2$ would have been unable to place differential demands on the subjects' behavior so as to confirm any predictions.

A more sophisticated alternative explanation for the results might involve Rosenberg's (1965) concept of evaluation apprehension. Perhaps certain groups of subjects--like white females or black males or students of low SES--were more likely to discuss the experiment among themselves because they were more fearful of being tested. Armed with prior knowledge of the research procedures, they may have resisted the experimenter's expectation manipulation as a way of telling him that they were aware of his efforts at deceiving them. This alternative does not, however, explain why gossip would be most likely to induce such resistance (a reactance effect) if a given subject's tester happened to be white rather than black, nor does it explain why the data for <u>all</u> groups of subjects--not just the most apprehensive--showed reactance effects in the gamelike atmosphere. If enough untested assumptions are included, evaluation apprehension could become a viable alternative explanation of the

data; at present, though, a motivational interpretation seems more parsimonious.

To the extent that reactance effects were observed in Experiments I and II, the findings appear to contradict those gathered in the "self-fulfilling prophecy" or "Pygmalion" paradigm initiated by Rosenthal and Jacobsen (1968). They found that when a teacher had been induced to have a high expectation of the abilities of certain randomly-selected students, the classroom performance of these students improved; by implication, a teacher's low expectation should debilitate performance. How can the results of the present research, in which an overtly-expressed low expectation seemed sometimes to motivate or challenge students to do their best on the WISC, be reconciled with those in the self-fulfilling prophecy tradition? The answer may lie in the word, overt. Chaikin, Sigler, and Derlega (1974) led undergraduate tutors to believe that a 10-year-old interviewee was either "quite bright" (IQ = 130) or "somewhat slow" (IQ = 85). It was found that tutors expecting a bright pupil leaned toward the interviewee, looked him in the eye, nodded their heads up and down, and smiled more frequently than tutors expecting a dull pupil; the former were also less likely to exhibit behaviors indicating dislike or disapproval, such as leaning backwards. Word, Zanna, and Cooper (1974) found that subjects exposed to an interviewer trained to emit standardized nonverbal cues of disapproval made a poorer impression on naive raters than those exposed to a nonverbally approving interviewer. So the subtle communication of a low expectation may indeed produce the well-known Pygmalion effect. However, an evaluator's low expectation may induce a poor performance on the part of examinees in such situations because it is so subtly expressed that any challenge to it is short-circuited by

the ambiguity in the situation. Research in which such things as atmosphere and expectation manipulations were either subtly or obviously communicated by the tester to the testee should serve to clarify the conditions under which one might anticipate a self-fulfilling prophecy rather than a reactance or challenge phenomenon.

The present research suggests that the variables of atmosphere and expectation may, when overtly expressed, interact with the subject's race and social class so as to have a considerable impact on his or her IQ score. If reliable and replicable, such findings would call into question Jensen's (1969) assertion that, since differences in the social and psychological environments to which white and black Americans are routinely exposed appear insufficient to account for interracial differences in mean IQ, a genetic explanation of these differences is called for.

Important questions remain, to be sure. Why, for instance, do interracial differences persist across parallel conditions? Even though high SES blacks performed remarkably well on the WISC when tested by a white experimenter in the evaluative atmosphere, why were they still outperformed by high SES whites in this same setting? Many answers are possible. The students were in the experimental situation for less than an hour; the cumulative effects of differential past experience for black and white subjects may not be so easily overcome. Furthermore, even though a white tester may, in general, have been more motivating than his black counterpart, he was probably not an unequivocally positive stimulus for a black student.

Since Sacramento is a medium-sized, highly mobile city in which the schools participating in the research were at most a few miles and in one instance a few blocks apart, it seems rather doubtful that the "high" and "low" categories created by the median split on the SES dimension

reflect substantially different gene pools. If so, if experience rather than heredity can be regarded as the major difference between the high and low SES groups, then the results would seem to imply that interracial differences in mean IQ can be erased or possibly even reversed depending on certain social-psychological characteristics of the test setting and the socio-economic background of the testee.

REFERENCES

Allen, S. A., Dubanoski, R. A., and Stevenson, H. W. Children's performance as a function of race of E, race of S, and type of verbal reinforcement. Journal of Experimental Child Psychology, 1966, 4, 248-256.

Baron, R. M., and Ganz, R. L. Effects of locus of control and type of feedback on the task performance of lower-class black children. Journal of Personality and Social Psychology, 1972, 21, 124-130.

Brehm, J. W. A theory of psychological reactance. New York: Academic Press, 1966.

Chaikin, A. L., Sigler, E., and Derlega, V. J. Nonverbal mediators of teacher expectancy effects. Journal of Personality and Social Psychology, 1974, 30, 144-149.

Crowne, D. P., and Marlowe, D. A new scale of social desirability independent of psychopathology. Journal of Consulting Psychology, 1960, 24, 349-354.

Doob, A. N., and Kirshenbaum, H. M. The effects on arousal of frustration and aggressive films. Journal of Experimental Social Psychology, 1973, 9, 57-65.

Duncan, O. D., Featherman, D. L., and Duncan, B. Socioeconomic background and achievement. New York: Seminar Press, 1972.

Garcia, J. IQ: The conspiracy. Psychology Today, 1972, 6 (9), 40ff.

Hays, W. L. Statistics for psychologists. New York: Holt, 1963.

Janis, I. L., and Field, P. B. Sex differences and personality factors related to persuasibility. In C. I. Hovland and I. L. Janis (Eds.) Personality and persuasibility. New Haven, Conn.: Yale University Press, 1959.

Jensen, A. R. How much can we boost IQ and scholastic achievement? Harvard Educational Review, 1969, 39, 1-123.

Katz, I. Experimental studies of Negro-white relationships. In L. Berkowitz (Ed.) Advances in experimental social psychology. Vol. 5. New York: Academic Press, 1970.

McClelland, L. Effects of interviewer-respondent race interactions on household interview measures of motivation and intelligence. Journal of Personality and Social Psychology, 1974, 29, 392-397.

Nowicki, S., and Strickland, B. R. A locus of control scale for children. Journal of Consulting and Clinical Psychology, 1973, 40, 148-154.

Nowlis, V. Research with the mood adjective checklist. In S. S. Tomkins and C. E. Izard (Eds.) Affect, cognition, and personality. New York: Springer, 1965.

Orne, M. T. On the social psychology of the psychological experiment: With particular reference to demand characteristics and their implications. American Psychologist, 1962, 17, 776-783.

Rosenberg, M. J. When dissonance fails: On eliminating evaluation apprehension from attitude measurement. Journal of Personality and Social Psychology, 1965, 1, 28-43.

Rosenthal, R. Experimenter effects in behavioral research. New York: Appleton-Century-Crofts, 1965.

Rosenthal, R., and Jacobsen, L. Pygmalion in the classroom. New York: Holt, 1968.

Rotter, J. B. Generalized expectancies for internal versus external control of reinforcement. Psychological Monographs, 1966, 80, Whole No. 609.

Sarason, I. G. The effect of anxiety and threat on the solution of a difficult task. Journal of Abnormal and Social Psychology, 1961, 62, 165-167.

Sattler, J. Racial 'experimenter effects' in experimentation, testing, interviewing, and psychotherapy. Psychological Bulletin, 1970, 73, 137-160.

Wechsler, D. WISC manual: Wechsler intelligence scale for children. New York: The Psychological Corporation, 1949.

Weiner, M., and Samuel, W. The effect of attributing internal arousal to an external source upon test anxiety and performance. Journal of Social Psychology, in press.

Wine, J. Test anxiety and the direction of attention. Psychological Bulletin, 1971, 76, 92-104.

Word, C. H., Zanna, M. P., and Cooper, J. H. The nonverbal mediation of self-fulfilling prophecies in interracial interaction. Journal of Experimental Social Psychology, 1974, 10, 109-120.

Table 1

Mean IQ
for Male and Female Subjects[a]

Sex	Atmosphere	Expectation	Black Subjects		White Subjects	
			Black Tester	White Tester	Black Tester	White Tester
Male	Evaluative	High	98.54	97.00	111.08	121.46
		Low	91.69	102.00	109.38	110.31
	Gamelike	High	93.23	96.46	107.15	109.46
		Low	96.23	105.62	109.15	118.46
Female	Evaluative	High	96.69	97.38	106.76	111.00
		Low	91.69	94.69	104.76	117.53
	Gamelike	High	89.30	101.08	103.15	112.30
		Low	97.69	97.46	110.15	115.92

[a]There were 13 subjects per cell. The following critical values for assessing the significance of differences between means have been derived from procedures for individual comparisons in Hays (1963):
17.09 ($p < .002$), 12.82 ($p < .02$), 10.83 ($p < .05$), 9.12 ($p < .10$).

Table 2

The Atmosphere x Expectation Interaction for Male and Female Subjects[a]

	Atmosphere	
Expectation	Evaluative	Gamelike
High	104.39	101.51
Low	102.75	106.33

[a]There were 104 subjects per cell.

Table 3

Personality Correlates of I.Q.

by Race and Sex of Subject

	Black Subjects			White Subjects		
Item	Combined	Male	Female	Combined	Male	Female
Aggression	-.32****	-.44***	-.19*	-.26****	-.35****	-.18*
Concentration	+.09	+.12	+.06	+.20***	+.31***	+.11
Egotism	-.20***	-.27***	-.14	-.12*	-.18*	-.08
Elation	-.14**	-.25***	-.02	-.10	-.04	-.18*
Fatigue	-.17***	-.24***	-.10	-.09	-.10	-.10
Sadness	-.26****	-.37****	-.20**	-.22****	-.13	-.30***
Skepticism	-.13*	-.07	-.20**	-.03	-.07	.00
Surgency	-.14***	-.18*	-.09	-.02	-.11	+.05
How Well (Self-rate)	+.08	+.06	+.11	+.18***	+.20**	+.16
Black is Beautiful[a]	+.16**	+.22**	+.07	-.05	-.09	+.03
General Anxiety[a]	-.26****	-.30****	-.19*	-.06	-.16	+.01
Test Anxiety[a]	-.18***	-.17*	-.18*	-.12*	-.12	-.10
I-E[a]	-.23****	-.25***	-.19*	-.33****	-.26***	-.39****

[a]These correlations were derived from items on the follow-up questionnaire, which a small number of subjects failed to complete.

****$p < .002$
***$p < .02$
**$p < .05$
*$p < .10$

Table 4

The Atmosphere x Expectation Interaction
for Male Subjects

	1972-73[a]		1973-74[a]		Combined[b]	
	Atmosphere		Atmosphere		Atmosphere	
Expectation	Evaluative	Gamelike	Evaluative	Gamelike	Evaluative	Gamelike
High	107.02	101.58	109.98	107.25	108.50	103.91
Low	103.35	107.36	104.98	108.15	104.16	107.75

[a]There were 52 subjects per cell.

[b]There were 104 subjects per cell.

Table 5

Mean IQ for Male Subjects High or Low in SES[a]

SES	Atmosphere	Expectation	Black Subjects: Black Tester	Black Subjects: White Tester	White Subjects: Black Tester	White Subjects: White Tester
High	Evaluative	High	99.62	114.60	113.72	126.76
		n =	8	10	18	17
		Low	93.22	107.67	111.25	115.36
		n =	9	9	16	14
	Gamelike	High	101.78	103.08	108.25	114.35
		n =	9	13	16	17
		Low	107.73	104.67	112.13	118.12
		n =	11	9	15	17
Low	Evaluative	High	97.22	98.12	108.75	110.22
		n =	18	16	8	9
		Low	95.47	100.53	99.60	108.50
		n =	17	17	10	12
	Gamelike	High	91.59	97.08	106.20	110.44
		n =	17	13	10	9
		Low	93.00	104.82	107.73	114.22
		n =	15	17	11	9

[a]The following critical values for assessing the significance of differences between means have been derived from procedures for individual comparisons in Hays (1963): 17.63 ($p < .002$), 12.84 ($p < .02$), 10.80, ($p < .05$), 9.04 ($p < .10$).

Table 6

Mean Anxiety and Aggression
for Male Subjects

	Anxiety[a]			Aggression[b]		
	Atmosphere			Atmosphere		
Expectation	Evaluative	Gamelike	p(diff)	Evaluative	Gamelike	p(diff)
High	3.24	2.45	.02	2.36	2.39	n.s.
Low	3.32	3.05	n.s.	3.19	2.89	n.s.
p(diff)	n.s.	.10		.05	n.s.	

[a] $F_{Atmosphere} = 5.74$, 1/412 df, $p < .05$

[b] $F_{Expectation} = 6.39$ 1/412 df, $p < .05$

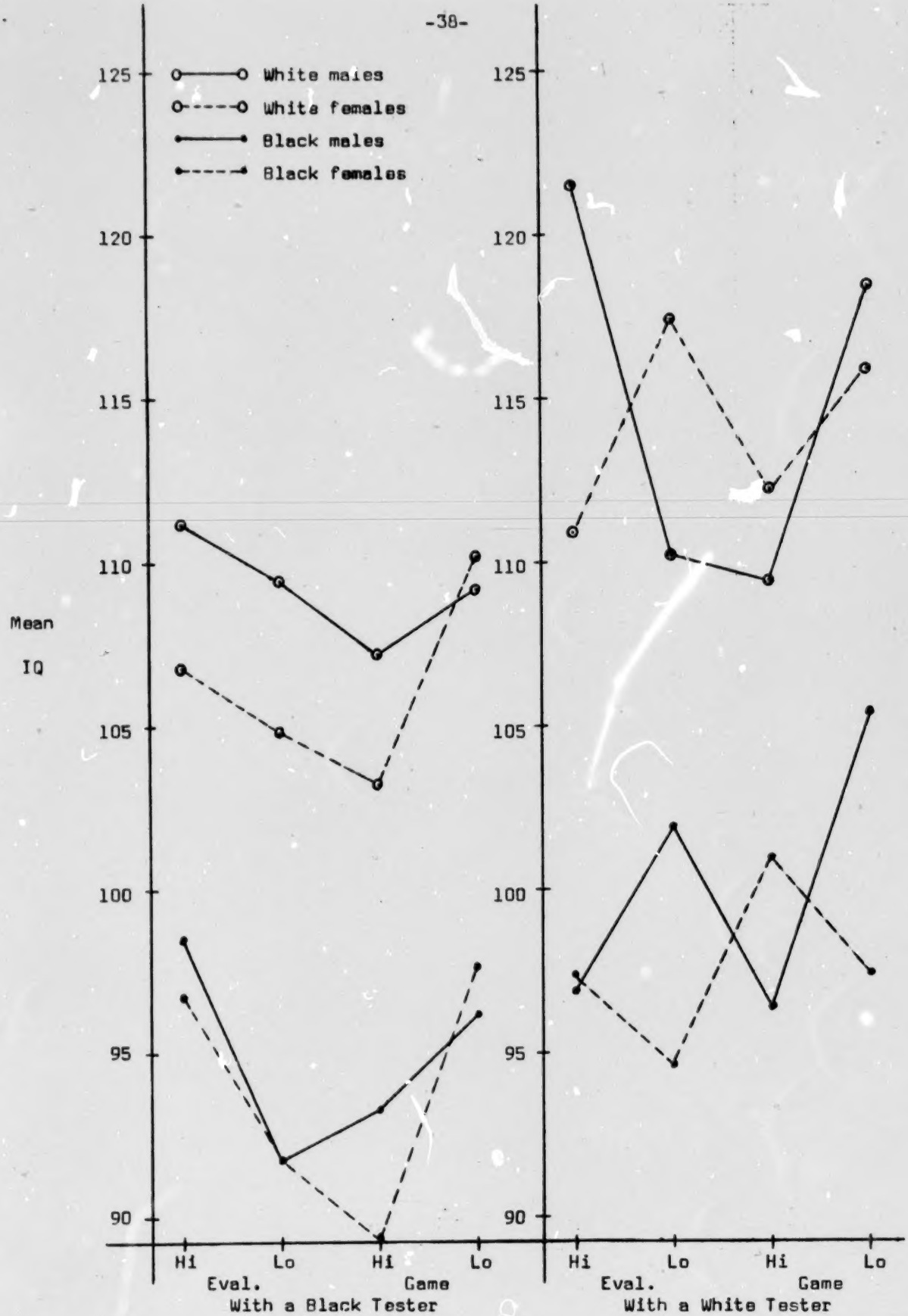

Fig. 1 -- Mean IQ for male and female subjects.

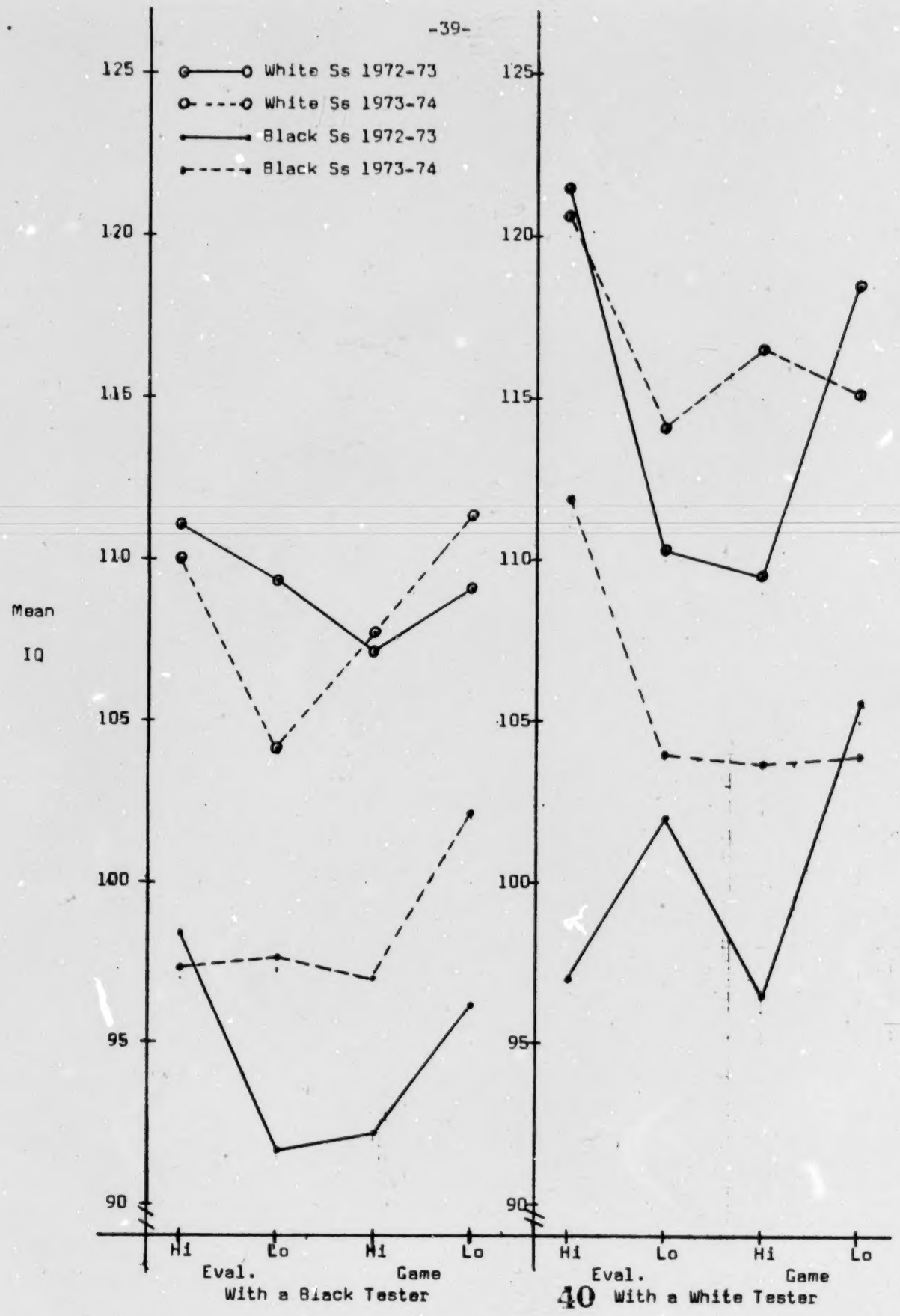

Fig. 2 -- Mean IQ for male Ss in the 1972-73 and 1973-74 experiments.

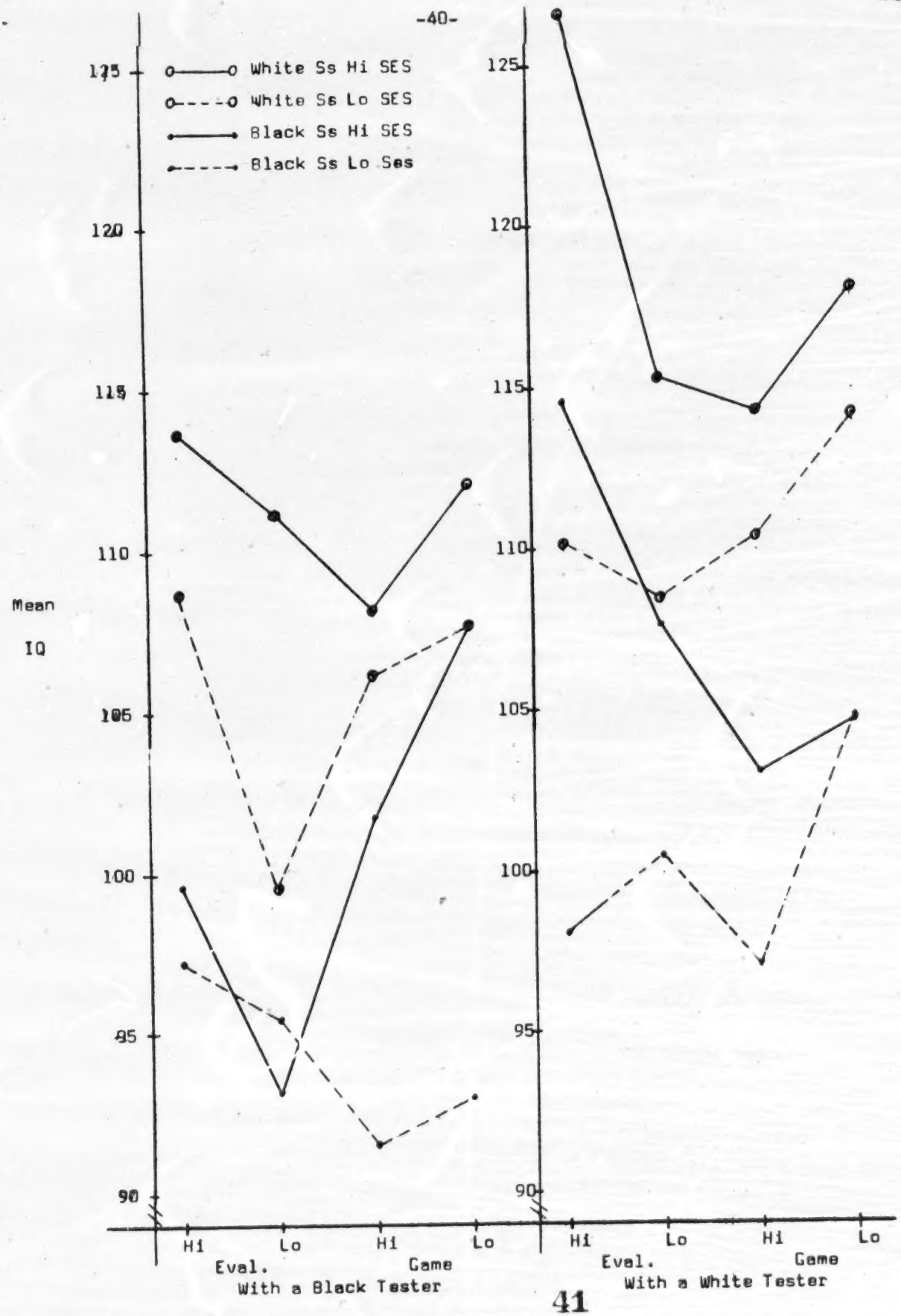

Fig. 3 -- Mean IQ for high and low SES male subjects.